



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2011

Modelling seasonal patterns in longitudinal profiles with correlated circular random walks

Riebler, A ; Held, L ; Rue, H

Abstract: Seasonal patterns, as they occur in time series of infectious disease surveillance counts, are frequently modelled using a superposition of sine and cosine functions. However, in some cases this might be too simple. We propose the use of circular second order random walks instead and extend this approach to multivariate time series of counts. A correlated Gaussian Markov random field approach combines a uniform correlation matrix with a circular random walk to allow the seasonal pattern to be similar across regions, say, but not identical. Thus, spatially-varying disease onsets may be accounted for. The methodology is applied to weekly number of deaths from influenza and pneumonia in nine major regions of the USA.

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-52612>

Conference or Workshop Item

Accepted Version

Originally published at:

Riebler, A; Held, L; Rue, H (2011). Modelling seasonal patterns in longitudinal profiles with correlated circular random walks. In: 26th International Workshop on Statistical Modelling, Valencia, 11 July 2011 - 15 July 2011, 503-508.

Modelling seasonal patterns in longitudinal profiles with correlated circular random walks

Andrea Riebler¹, Leonhard Held¹ and Håvard Rue²

¹ Division of Biostatistics, Institute of Social and Preventive Medicine, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland;

Email: andrea.riebler@ifspm.uzh.ch, leonhard.held@ifspm.uzh.ch

² Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway; Email: havard.rue@math.ntnu.no

Abstract: Seasonal patterns, as they occur in time series of infectious disease surveillance counts, are frequently modelled using a superposition of sine and cosine functions. However, in some cases this might be too simple. We propose the use of circular second order random walks instead and extend this approach to multivariate time series of counts. A correlated Gaussian Markov random field approach combines a uniform correlation matrix with a circular random walk to allow the seasonal pattern to be similar across regions, say, but not identical. Thus, spatially-varying disease onsets may be accounted for. The methodology is applied to weekly number of deaths from influenza and pneumonia in nine major regions of the USA.

Keywords: circular random walk; infectious disease surveillance; INLA; Kronecker product; multivariate time series of counts.

1 Introduction

Time-series of infectious disease counts are marked by occasional outbreaks, but furthermore there are frequently seasonal variations, for instance harder strikes in winter than summer. To model seasonal variation a superposition of sine and cosine functions is often used, where the amplitudes can be described by a fixed coefficient or, to be more flexible, by smoothly time-varying coefficients, see for example Harvey and Koopman (1993), Eilers et al. (2008), Paul et al. (2008) or Fanshawe et al. (2008). However, in some cases this approach might be too simplistic and specific seasonal variations, for example sharp peaks around Christmas, might not be captured (Harvey and Koopman, 1993). Circular random walks (CRWs) are similar in spirit to periodic splines (see Harvey and Koopman, 1993) and represent a flexible alternative to adequately capture seasonal variations. In a multivariate setting, where different regions, say, show a similar seasonal pattern which is, however, likely to vary across regions, we propose the use of correlated CRWs. Analyses are performed using integrated nested

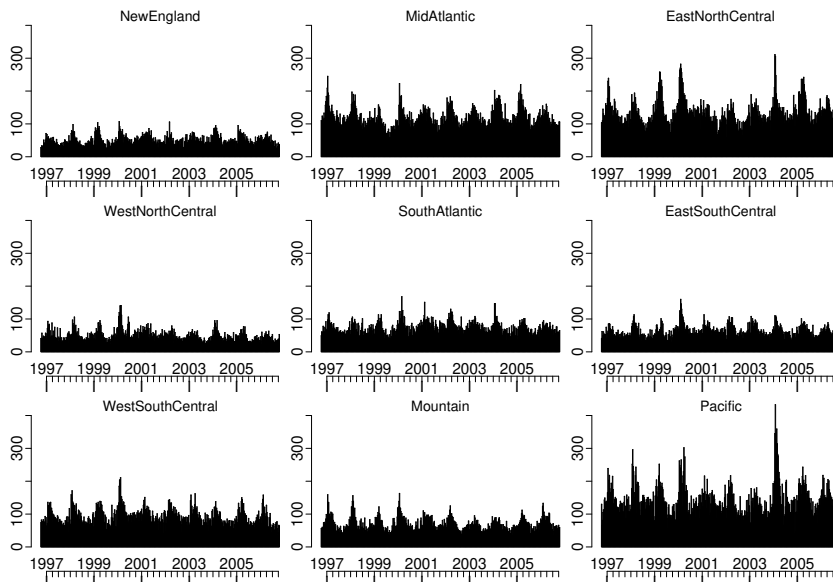


FIGURE 1. Weekly number of deaths from influenza and pneumonia in the USA from 40/1996 - 39/2006.

Laplace approximations (INLAs), see www.r-inla.org, which is a fast deterministic alternative to MCMC for latent Gaussian random field models (Rue et al., 2009). We apply the methodology to weekly number of deaths from influenza and pneumonia in the USA, previously analysed by Paul et al. (2008). Using the deviance information criterion (DIC) we compare the correlated approach with a model using independent CRWs for each region and a model assuming a common seasonal pattern across all regions.

2 Weekly data of influenza in nine regions of the USA

Weekly data on the number of deaths from influenza and pneumonia are provided for the weeks 40/1996 to 39/2006 in nine major geographic regions of the USA, see Figure 1. Region-specific population counts are not available for all years. Thus, we used the population counts derived from a census in the year 2000 in our analysis.

Let y_{tr} denote the number of deaths at time point t in region r , $r = 1, \dots, R$, with $R = 9$. In our application, time is divided into weeks from 40/1996 to 39/2006, so that $t = 1, \dots, 520$. We adopt a Poisson model with mean $n_r \lambda_{tr}$, where n_r denotes the population counts in region r (in the year

2000). To adequately model the seasonal pattern in a general and flexible way we use a CRW of second order (CRW2) for the 52 weeks. The precision matrix of a CRW2 is given by

$$\mathbf{R}^{CRW2} = \kappa \begin{pmatrix} 6 & -4 & 1 & 0 & \cdots & 0 & 1 & -4 \\ -4 & 6 & -4 & 1 & 0 & \cdots & 0 & 1 \\ 1 & -4 & 6 & -4 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -4 & 6 & -4 & 1 \\ 1 & 0 & \cdots & 0 & 1 & -4 & 6 & -4 \\ -4 & 1 & 0 & \cdots & 0 & 1 & -4 & 6 \end{pmatrix}, \quad (1)$$

with unknown precision parameter κ . As for all circulant matrices, only one column or row is sufficient to derive the whole structure matrix (Rue and Held, 2005, Section 2.6.1). To allow for similar but not equal seasonal patterns across the nine regions, we correlate the single CRW2s using the precision matrix $\mathbf{P} = \mathbf{C}^{-1} \otimes \mathbf{R}^{CRW2}$. Here, \mathbf{C}^{-1} is the inverse of a 9×9 uniform correlation matrix $\mathbf{C} = (1 - \rho)\mathbf{I} + \rho\mathbf{J}$, where ρ denotes the unknown correlation parameter, \mathbf{I} the identity matrix, \mathbf{J} a matrix of ones, and \mathbf{R}^{CRW2} is the precision matrix given in (1). In addition to seasonal variation, the disease incidence, as displayed in Figure 1, shows occasional outbreaks. To address such temporal dependence beyond seasonal variation, we additionally introduce an autoregressive process of order 1 (AR1) again coupled with a uniform correlation matrix. The linear predictor follows as:

$$\log(\lambda_{tr}) = \mu_r + \alpha_{tr} + \beta_{(t \bmod 52)r}, \quad (2)$$

where μ_r denotes the region-specific intercept, α_{tr} the outbreak-specific component modelled as a correlated AR1 and $\beta_{(t \bmod 52)r}$ the seasonal component modelled as a correlated CRW2.

All 5 hyperparameters (the seasonal precision, the correlation between the CRWs, precision and autoregressive parameter of the AR1 processes and correlation between the AR1s) are treated as unknown. For the unknown precision parameters we use gamma hyper-priors, namely a $Ga(1, 0.00005)$ for the precision κ of the correlated random walk, and a $Ga(0.1, 0.001)$ for the precision of the AR(1) process as proposed by Schrödle et al. (2011). For the Fisher's z-transformed autoregressive parameter we use a normal distribution with zero-mean and variance 0.2^{-1} , corresponding to a U-shaped prior. The same prior is used for the transformed correlation parameters between the CRWs and the AR1s. Here, the general Fisher's z-transformation (Fisher, 1958, page 219) is used, which ensures that the correlations only take values between $(-1/(R-1), 1)$, so that \mathbf{C} is positive definite without imposing an additional constraint, see also Riebler et al. (2011).

TABLE 1. DIC for three different models using a CRW2 to model seasonal variation in the nine major regions of the USA.

	common CRW2	region-specific CRW2 uncorrelated	region-specific CRW2 correlated
DIC	36707	36716	36704

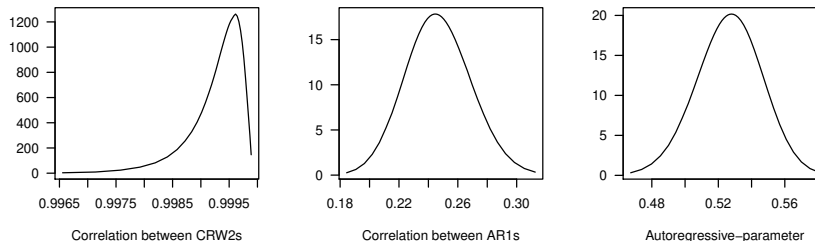


FIGURE 2. Approximated posterior marginals for the correlation parameters and the autoregressive parameter.

3 Results

We compared the model, which uses correlated CRW2s to model the seasonal pattern in the nine regions, with a model assuming independent CRW2s and a model assuming a common CRW2. The DIC values for all three models are shown in Table 1. The model assuming correlated region-specific CRW2s is classified as the best model for which Figure 2 shows the approximate posterior marginals for both correlation parameters and the autoregressive parameter. The correlation between the seasonal components is close to unity (0.999; 95% CI: [0.998, 1]). Figure 3 shows the seasonal pattern (mean within 95% CI) for NewEngland, and for the other regions the pair-wise differences of the estimated mean seasonal effects to NewEngland are shown. For all regions, the seasonal component is higher in the winter months and lower during the summer, but some small differences occur across regions. For example, in Mountain the peak in the winter months is higher, while the pattern in summer is lower compared to NewEngland. In SouthAtlantic it is the other way around.

Of note, the seasonal pattern is not completely smooth. The decreasing effect at the end of the year and the increasing effect at the beginning might be explained by a Christmas effect, where few cases are reported around Christmas but many after the holidays. Peaks throughout the year are not completely clear and need to be investigated in detail.

Turning to the correlated AR1 processes, we note that the estimated autoregressive parameter is 0.53 (95% CI: [0.49, 0.56]) and the correlation

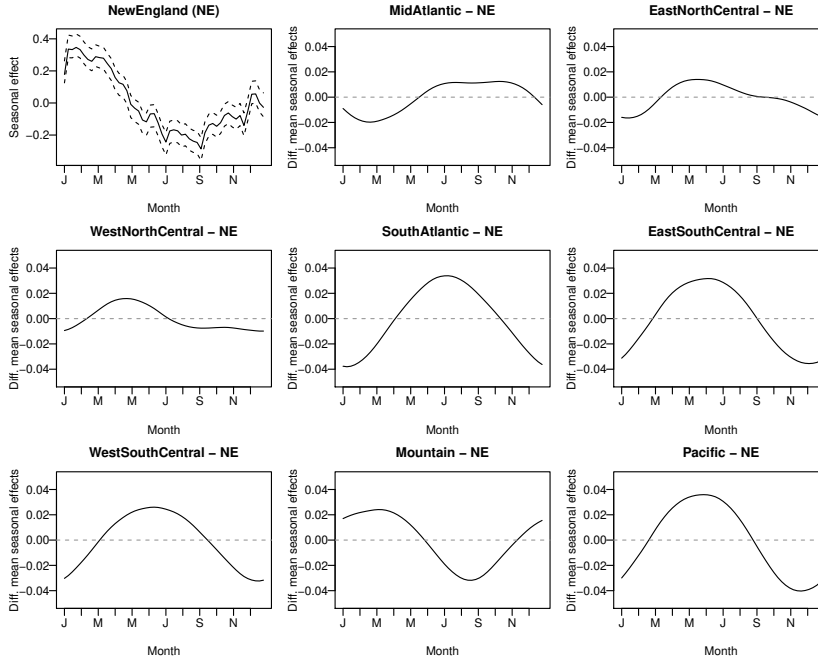


FIGURE 3. Estimated seasonal effects (mean and 95% CI) for NewEngland (NE) (top left). For the other regions the difference of the mean seasonal effects to NE is shown.

between the processes is estimated to be 0.25 (95% CI: [0.20, 0.29]) and thus also clearly different from zero.

4 Discussion and outlook

We proposed the use of correlated CRW2s for modelling seasonal variation in multivariate time series of counts. We applied the methodology to weekly numbers of deaths from influenza and pneumonia in nine major regions of the USA. Although, the correlation between the single seasonal trends was close to unity, this model was preferred compared to a model with one common seasonal component.

In certain aspects the CRW2 represents a quite flexible approach, as the seasonal pattern is not restricted in its functional form, so that also sharp peaks can be captured. However, it assumes that the temporal pattern repeats every 52 weeks, whereas ideally we would like to account for time-varying disease onsets.

The modulation model proposed by Eilers et al. (2008) is more flexible in this aspect. However, here the (co)sine function might be too simple in

certain applications. An unstructured non-parametric model as defined in Rue and Held (2005, page 122f) can also account for time-varying disease onsets. However, here the week indicators are treated exchangeable so that the seasonal pattern is not required to be smooth. Both the modulation model of Eilers et al. (2008) and the seasonal model of Rue and Held (2005, page 122f) can be implemented in INLA and could also be coupled across regions using a uniform correlation matrix. Currently, we are working on a comparison and if possible a combination of these models. Furthermore, we are exploring possibilities to include spatial correlation between the nine geographical regions of the USA.

Acknowledgments: This work received support from the Swiss National Science Foundation and the Research Council of Norway.

References

- Eilers, P.H.C., Gampe, J., Marx, B.D. and Rau, R. (2008). Modulation models for seasonal time series and incidence tables. *Statistics in Medicine*, **27**, 3430–3441.
- Fanshawe, T.R., Diggle, P.J., Rushton, S., Sanderson, R., et al. (2008). Modelling spatio-temporal variation in exposure to particulate matter: a two-stage approach *Environmetrics*, **19**, 549–566.
- Harvey, A. and Koopman, S.J. (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association*, **88**, 1228–1236.
- Paul, M., Held, L. and Toschke, M.A. (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine*, **27**, 6250–6267.
- Riebler, A., Held, L. and Rue, H. (2011). Correlated multivariate age-period-cohort models. Technical report. University of Zurich. Available from: <http://www.biostat.uzh.ch/research/manuscripts/cmapc.pdf>.
- Rue, H., and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman & Hall/CRC Press.
- Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, **71**, 319–392.
- Schrödle, B., Held, L. and Rue, H. (2011). Assessing the impact of network data of the spatio-temporal spread of infectious diseases. Technical report. University of Zurich. Available from: <http://www.biostat.uzh.ch/research/manuscripts/schroedle.etal.2011.pdf>.